

The Structured Process Modeling Method (SPMM)

What is the best way for me to construct a process model?

Jan Claes^{1*}, Irene Vanderfeesten², Frederik Gailly¹, Paul Grefen², and Geert Poels¹

¹ Ghent University, Department of Business Informatics and Operations Management,

Tweekerkenstraat 2, 9000 Ghent, Belgium

² Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences,

Postbus 513, 5600 MB Eindhoven, The Netherlands

* Corresponding author. Tel.: +32 9 264 34 17; fax: +32 9 264 42 86.

E-mail addresses: jan.claes@ugent.be (J. Claes), i.t.p.vanderfeesten@tue.nl (I. Vanderfeesten), frederik.gailly@ugent.be (F. Gailly), p.w.p.j.grefen@tue.nl (P. Grefen), geert.poels@ugent.be (G. Poels).

Abstract. More and more organizations turn to the construction of process models to support strategical and operational tasks. At the same time, reports indicate quality issues for a considerable part of these models, caused by modeling errors. Therefore, the research described in this paper investigates the development of a practical method to determine and train an optimal process modeling strategy that aims to decrease the number of cognitive errors made during modeling. Such cognitive errors originate in inadequate cognitive processing caused by the inherent complexity of constructing process models. The method helps modelers to derive their personal cognitive profile and the related optimal cognitive strategy that minimizes these cognitive failures. The contribution of the research consists of the conceptual method and an automated modeling strategy selection and training instrument. These two artefacts are positively evaluated by a laboratory experiment covering multiple modeling sessions and involving a total of 149 master students at Ghent University.

Keywords: modeling support, smart business process management, cognitive aspects of modeling, process of process modeling, process model quality.

1 Introduction

In today's competitive markets with challenges in terms of globalization, mass-customization and risk control, it is considered important for organizations to manage and control their business

processes thoroughly. Therefore, many organizations nowadays spend a great deal of effort to build and maintain a collection of business process models (or “process models” for short). These models represent various aspects of the business processes, such as the control, communication and information flows, while abstracting from individual process instances. The process models are used to support a diversity of process management tasks ranging from the strategical to the operational level: communication, documentation, analysis, (re)design, simulation, execution, etc. [1,2].

Unfortunately, regardless of their importance and potential, case studies report many issues with the quality of process models in organizations [3–5]. Hence, researchers have built a large body of knowledge about the quality of conceptual (process) models [6–21]. Nevertheless, research about operational guidance on how to create high quality process models appears to be limited to the development of general guidelines about the ideal properties of produced models [22–25] and to the spontaneous description of best practices that emerged among process modeling experts [26,27]. Thus, there is a lack of sound operational support to help a modeler in translating her/his mental image of the real-world business process into a high quality process model [21].

Considering on the one hand the importance of process modeling and on the other hand the reported quality issues and lack of operational support, the research objective addressed in this paper is to develop a practical method that helps modelers to implement an optimal process modeling strategy. Disregarding those quality issues that are related to a modeler’s imperfect knowledge, the research question is: How can the cognitive processing of the modeler be supported to create process models effectively and efficiently, given her/his current knowledge about the process and about modeling?

We believe that this cognitive support is optimal if it is aware of and adapts to the characteristics and behavior of the user. We define such differentiated and adaptive support as “smart support”, similar to the definition of Smart Technologies, which are technologies that are aware of and adapt to (changes in) their situation [28]. Hence, this paper presents smart support in the context of business process modeling. More concrete, a smart conceptual method was developed that assists modelers in discovering and training their individual optimal process modeling strategy: the Structured Process

Modeling Method (SPMM). Further, a prototype tool implementation was developed that enables a modeler to autonomously execute the method. This implementation is smart in the sense that it measures the cognitive profile of the modeler to decide which modeling guidelines it proposes and that it adapts the information offered to the user based on her/his previous interactions.

Both the conceptual method and the supporting implementation were tested via an extensive lab experiment. The results indicate (i) that it is possible to deliberately adjust one's modeling behavior with a limited, unsupervised intervention, (ii) that this adjustment by the method indeed has a substantial beneficial effect on process modeling, and (iii) that the users perceive the implementation as useful to improve their modeling efficacy. From the performed lab experiment, it appears that the method in its current form mainly helps reducing the effort and duration of process modeling (efficiency of modeling), without significantly impacting the end quality of the resulting process model (effectiveness of modeling).

This paper is structured as follows. Section 2 elaborates on the related research and the theoretical background. Section 3 presents the developed method and its implementation. Section 4 describes the evaluation of the research with a large-scale experiment. Section 5 discusses the potential impact and the limitations of this work. Section 6 concludes with a summary of the paper and of future research.

2 Preliminaries

This section discusses related work (Section 2.1) and the theoretical background (Section 2.2).

2.1 Related work

Literature on (business) process model quality has mainly focused on the specification of quality frameworks, the identification of quality dimensions and the development of quality measures. A brief overview is presented hereafter. The LSS framework of Lindland, et al. compares the goals and means of conceptual models to define different quality dimensions and their mutual relationships [6]. Inspired by linguistic concepts, they identify syntactic, semantic and pragmatic quality as the main model quality dimensions. Nelson, et al. present a conceptual modeling quality framework that extends the

LSS framework with 6 more dimensions [7]. These dimensions are derived from insights of the BWW framework [29], which additionally focuses on the dynamic aspects of conceptual modeling. Rockwell and Bajaj further extended this process-oriented view on modeling quality by investigating the cognitive aspects of the efficacy of the model creation and model understanding processes [8]. Further, in the more specific context of process modeling, Krogstie, et al. have further improved the dynamic aspect of the semantic and pragmatic quality components of the LSS framework [9] and Becker, et al. have specified process modeling quality dimensions at a more practical, operational level [10].

The different dimensions of these quality frameworks are studied in more detail. In the context of process models, the research has focused mainly on the quality dimensions of model correctness [11] (related to syntactic and semantic quality), model understandability [12–14] and adequacy [15] (related to pragmatic quality) and model maintainability [13,14] (related to modeling efficiency).

Further, a large number of metrics are defined that describe these dimensions directly or that are used as indirect indicators for these dimensions. For example, researchers have related various metrics of process model complexity to the model correctness [16–18], understandability [12,13] and maintainability [13]. According to Laue and Mendling, another indicator for the model correctness is the structuredness of the model [11]. Likewise, the model understanding is also linked to the modeling language selection [19] and to the formulation of activity labels [20]. More examples can be found in the extensive and recent literature reviews about process model quality of Sánchez-González, et al. [14] and Moreno-Montes de Oca, et al. [21].

Based on the literature about process model quality, operational support was developed for creating high quality process models. In Seven Process Modeling Guidelines (7PMG), Mendling, et al. describe seven guidelines of good modeling practice [22]. Examples are “*Use one start and one end event*”, “*Avoid OR routing elements*”, and “*Decompose the model if it has more than 50 elements*” [22], p. 130. The 7PMG have been complemented and specified with concrete thresholds in [14] and [23]. Similarly, in Guidelines of Modeling (GoM), Becker, et al. define a set of guidelines, which describe desired properties of a constructed process model, such as correctness, relevance and clarity

[10]. In [24] and [25] La Rosa, et al. explain how models can be made more understandable by optimizing their internal structure and aesthetics respectively. Examples are modularization, block-structuring, highlighting and pictorial annotation. These guidelines all offer relevant and valuable support for modelers, but they have been criticized to be incomplete [21], to be too abstract [21,22] and/or to lack empirical evaluation [30,31]. This illustrates the need for a more concrete process modeling support that is tailored to its users (i.e., the need for smart process modeling support).

Inspiration for such smart process modeling support can be drawn from other research domains. The field of personalized teaching material, for example, addresses similar goals and has already produced interesting knowledge about smart support [32]. In the context of adaptive learning and differentiated instruction, the existing literature focuses heavily on computer assisted instruction systems (CAI). These systems - often called ‘adaptive hypermedia systems’ - adapt the pace, order and representation of information flows to various characteristics of the user [33,34]. They often have a layered architecture consisting of a Domain Model (containing the domain ontology and a learning goal hierarchy), User Model (containing the user knowledge base and user characteristics) and Adaptation Model (containing content selection rules) [35]. One particular example worth mentioning is the two-source adaptive learning system of Tseng, et al., which is showed to improve learning achievements of high school students [36]. Just as the implementation presented further in this paper, their system adapts to the user’s individual learning style and to their learning behavior. In general, it can be noticed how most of the presented solutions use similar cognitive concepts as the ones discussed in this paper to adapt the support to the users [37,38]. The main difference with the approach presented in this paper is that in the learning context the selection of content is not adapted to the user. Whereas in a learning context it is the goal that all learners acquire the same knowledge, in a modeling context they only need to learn what is necessary to apply their individual optimal modeling strategy.

2.2 Theoretical background

In order to address the identified research gap, we started a methodological research line in 2011 to develop this practical process modeling support method. Fig. 1 shows the different steps and resulting publications of this methodological research flow. Based on an extensive collection of

recorded modeling instances, the research started with the documentation of various process modeling patterns and of potential relations with the resulting model quality [39,40]. Research accelerated after the development of the PPMChart, a research instrument that visually represents the tool operations of a modeler while creating a single process model [41].

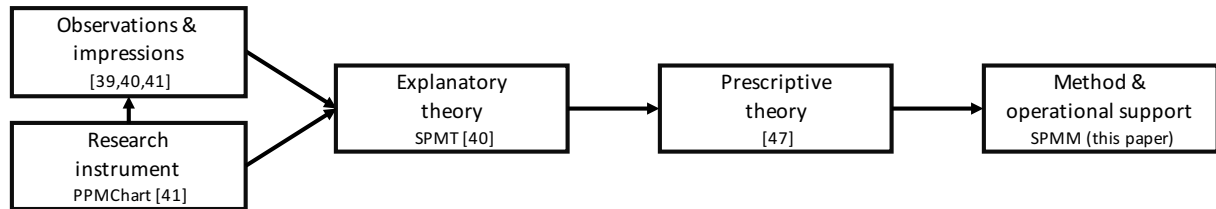


Fig. 1. Steps and papers leading to the current research contribution

Next, three particular cognitive process modeling techniques were identified [40]. *Flow-oriented process modeling* is a technique where the modeler builds the model in consecutive parts that (s)he creates, edits and completes before turning to the next one. *Aspect-oriented process modeling* is a technique where the modeler builds the model in multiple iterations, in each going through the whole model, but working on only one aspect (or a few related aspects) of the model, such as the creation of elements, the structure, the lay-out, etc. *Combined process modeling* is a technique where the modeler combines the previous ones: (s)he creates the model in a flow-oriented manner, occasionally pausing to improve a specific aspect of the partial model so far.

Based on these observations a theory was developed that explains the described relations between a user's modeling approach and the quality of the produced process model: the Structured Process Modeling Theory (SPMT) [40]. It is based on the Cognitive Load Theory, which states that cognitive overload causes us to make mistakes and to think slower [42]. In order to avoid cognitive overload people apply a technique that is called cognitive sequencing [43]. This means that information is processed relatively more sequentially instead of simultaneously. According to the SPMT, cognitive overload can be minimized during process modeling if the cognitive sequencing techniques are applied in a structured way that fits with a modeler's cognitive preferences [40].

The SPMT describes three cognitive preferences of a modeler that determine how well a process modeling strategy fits. A person's *learning style* cognitive preference describes how (s)he takes in new

information. The concept encompasses different dimensions, but we consider only the *understanding* dimension, which makes a distinction between relatively sequential and relatively global learners. Sequential learners process information in a linear way, step-by-step, in a steady progression and in connected chunks, whereas global learners jump directly to difficult material and progress in intuitive leaps [44]. The *field dependency* cognitive preference indicates how much one relies on details for understanding material. Field-dependent people prefer specific information and have a short attention span, whereas field-independent people are better at abstract reasoning and can work more focused [45]. The *need for structure* cognitive preference indicates how much people desire structure in new material and how they react to missing structure [46].

Being an explanatory theory, the SPMT defines potential causes for the observed phenomena, but it has no predictive power. In order to progress the theory towards a prescriptive theory, in [47] the relative importance of each relation is investigated and the necessary concepts are operationalized. More in detail, the theory of [47] proposes to first match a modeler's learning style with one of the identified sequencing techniques: sequential learners should apply the flow-oriented style, global learners are matched with the aspect-oriented style and people in between are instructed to apply the combined style. Next, the field dependency score is used to provide additional guidelines: field-dependent modelers should take frequent short breaks, work on smaller parts at once, model in the provided order and keep all parts of the model connected during modeling, whereas field-independent modelers should not be afraid to deviate from these guidelines. The need for structure influences the impact of the above instructions and is thus used to set the right expectations for the modeler.

In summary, the description of observations in [39,40] forms a descriptive theory, from which we developed the explanatory theory SPMT [40], which then evolved to the prescriptive theory proposed in [47]. We refer to the work of Gregor to learn more about different types of theory and their cumulative development [48]. This paper describes the next research step: based on the knowledge expressed in the developed theories, a concrete method and the related smart implementation is proposed. Being design science research, which addresses the “how” dimension, this research differs from the previous work which is behavioral research about the “what” dimension. In other words,

whereas the previous studies focus on answering questions in order to build the required knowledge, this study finally focuses on solving a problem by the development and the investigation of artefacts (based on and extending the previously developed knowledge).

3 The Structured Process Modeling Method (SPMM)

This section describes the development of the conceptual method and of the accompanying implementation of the prototype operationalizing the method. Based on the preparatory research described in the theoretical background section, it may seem obvious how to define the method to support process modelers in selecting and training their optimal modeling style. Indeed, the method is no more than the execution of the next three steps: (i) measure cognitive preferences, (ii) determine the matching process modeling strategy and (iii) train this process modeling strategy. Nevertheless, the translation of these steps into executable instructions is not trivial. For the first step, suitable metrics for the identified cognitive variables need to be found. The second step is the application of the prescriptive theory of [47]. Finally, the third step is the most challenging because an optimal training strategy needs to be developed. Whereas the second step is fully prepared by previous research, the first and third step needed further investigation which is described in Sections 3.1 and 3.2 respectively.

3.1 Measure the cognitive profile

Validated measurements were searched for the cognitive variables of the Structured Process Modeling Theory (the three cognitive preferences discussed in Section 2.2). This subsection reports on the design choices that were made while selecting and adapting these measurements.

Concerning the sequential/global learning style, different measurements exist. The *Spy Ring History* and the *Smugglers* tests of Pask have been investigated thoroughly and these measures are deemed to be reliable and valid [49]. However, these tests are complex, demanding and lengthy (about 31 hours each) [37,49]. Therefore, they are not considered further for this research. Then, two other measures were found that are used widely in practice: the Index of Learning Styles by Felder & Silverman [44] and the holist/serialist measure by Ford [50]. To the best of our knowledge, the validity of these measures is not known. Based on the higher number of citations, the measure of Felder and

Silverman was selected for use in this work. This survey measures five dimensions of learning style of which we are only interested in the sequential/global learner dimension. Therefore, we omitted the irrelevant items resulting in an 11-item questionnaire, presented in Appendix A. Application of the measurement results in an odd integer score between -11 (global learner) and +11 (sequential learner).

Traditionally, the field dependency is measured by the Rod-and-Frame Test in which a participant has to try to position a rod perfectly vertical while a tilted frame in the background may be hindering this task depending on the participant's field dependency [45]. In order to facilitate the reproduction of the measurement on a larger scale, an alternative test on paper was developed: the embedded figures test in which the participant needs to find a simple line pattern within a more complex pattern of lines. Different variations of this test exist. The Group Embedded Figures Test is the original variant, of which validity and reliability have been showed to be strong [51]. Because this test is not widely available, we turned to another commonly used variant: the Hidden Figures Test (HFT) [52]. This variant is also considered valid and reliable [52]. It is presented in Appendix B. The test results in a real score between 0 (field-independent) and 1(field-dependent).

The need for structure is generally measured with the Personal Need for Structure scale of Thomson, et al. [53]. This questionnaire has shown to be valid and reliable [46,53]. It is applied according to the instructions of [46], as presented in Appendix C. Note how the variable is based on two concepts: desire for structure and reaction to missing structure. Only the first concept has to be measured for the next step of the method, but both concepts are required to evaluate the method and its implementation. Hence, the full scale is currently used for the measurement, but only the results of the relevant sub-concept are included in the first step of the method. After scoring the answers of a test subject, an integer score between 1 (low desire/reaction) and 6 (high desire/reaction) is obtained.

The measurement instruments for the three variables were integrated in a single, digital cognitive assessment implementation [54]. This required to make (further) slight adaptations to the measurement protocol. The learning style questionnaire was shortened to only include the relevant questions. Recall that the sequential versus global distinction used in the method forms only one dimension of learning

style. Next, instead of presenting two sheets of paper each containing eight assignments for the Hidden Figures Test, the digital version presented the assignments one by one. However, the participants were allowed to browse through the assignment in both directions (within the normal time limit). Apart from the medium, the selection of learning style questions and the consecutiveness of HFT assignments, no changes were made to the validated measurements. In theory, these adaptations can have influenced their documented validity, but the effect of the changes is assumed to be minimal.

3.2 Train the selected process modeling strategy

After having measured the three relevant cognitive variables and having selected an optimal process modeling strategy according to the prescriptive theory of [47], the selected modeling strategy has to be trained. In order to develop a smart training instrument (i.e., adapting to the cognitive profile and to the behavior of the user), next design decisions were made.

Differentiated. Because different cognitive preferences ask for different process modeling strategies, the smart instrument should adapt its contents to the user. 12 different general workflows are included in the training instrument. They correspond to the 12 possible combinations of directives.

Digital. To increase the smart aspect of the instrument, it should not only be differentiated (in this case adapting to the user), but also adaptive (adapting to the behavior of the user). Therefore, a digital instrument is preferred (i.e., a ‘tutorial’). This way, adaptations can be pre-coded precisely and can automatically be performed during the training session. There are a number of additional advantages of developing a digital tutorial in the context of this research. In contrast to a lecture-based training, there is no confusion about the particular contents of the training, because every detail is documented in (the program code of) the tutorial itself. Further, it can be exactly reproduced in multiple situations, which is preferred for between-subject studies. Next, with a digital instrument it is feasible to perform a simultaneous, yet distributed training session. A supervised environment may be recommended for some audiences though, in order to avoid distractions. On the other hand, one major challenge lies exactly in the absence of a human administrator who may be able to adapt the training more easily to the audience and to unforeseen questions and circumstances.

Short. The difficulty of changing someone's behavior should not be underestimated. The perfect training instrument leaves room for repetition and rehearsal, multiple iterations of knowledge acquisition and knowledge processing, hands-on practice, etc. Therefore, one may expect a multi-day intensive training is required to learn the selected process modeling strategy with all its nuances and subtleties. Yet, because a digital approach was preferred, the goal was set to develop a short but effective intervention. The intended duration was set to the arbitrary value of one hour only.

Repetitive. During the one-hour training, the user needs to process much information. Hence, it is recommended to provide the information in cumulative portions. This formula facilitates the continuous repetition and extension of knowledge from different viewpoints, which also suits the suggested iterative approach of alternating between knowledge acquisition and knowledge processing. At the end of the tutorial, a summary of the key information is presented as a final repetition.

Interactive. To create and increase adaptation opportunities, the tutorial has to be really interactive, which can be achieved by introducing practical exercises and challenging questions. They are also ideal to support the knowledge processing phases. However, it was decided to not make use of (longer-lasting) practical exercises because of the time limit that was set. This was compensated by varying between various types of questions: open and closed, knowledge and insight related, objective and subjective, general and specific.

Feedback. Considerable effort was spent to provide adaptive feedback on the question responses. For the multiple-choice questions, targeted feedback was formulated for every possible answer. For right answers, a confirmation and explanation message was provided. Wrong answers were selected to deal with potential confusions and thus the feedback for these answers tries to solve the confusion. As such, 49 different feedback statements were prepared.

Reflection. Providing feedback helps each user to reflect on the material and her/his individual progress. In order to further increase differentiated reflection opportunities, after each part (i.e., learning style, field-dependency, desire for structure), the user was asked explicitly to reflect on what was learned so far and to formulate a couple of key points and goals for her/his future modeling

assignments. An overview of these answers was presented again to the user in the concluding section where the contents of the tutorial were summarized in two consecutive screens.

Adaptive. Because of the absence of a human operator and for practical reasons, it is challenging to build a digital tutorial of this complexity that adapts to the behavior of the user. Yet, a further opportunity for adaptiveness was identified in regulating the amount of information that is offered to the user. Some users need more explanation than others, because they do not discover certain nuances in the material themselves. Therefore, these nuances are initially not explained. Instead, a number of the questions are directed to see if the user understands nuances that were not discussed before. They are formulated in such a way that they try to trick the user to answer questions incorrectly, but without confusing her/him, after which the feedback can resolve the misconception. This is also mentioned explicitly to the user in a pop-up message that appears after three incorrect answers are given in the tutorial. The message reassures not to worry about wrong answers, because the questions are there to clarify nuances that were not even explained before, which in this way is only presented to whoever didn't discover this her/himself. Furthermore, by letting the users re-answer the multiple-choice questions until they find the right answer, they can (un)consciously regulate the amount and pace of feedback. However, as soon as the correct answer is given, the feedback for all the (other) wrong answers is presented to the user who can then decide if (s)he want to read it or not.

Tolerant for inventions. Humans are difficult to train. Because of a mixture of conservatism, fit with existing knowledge, overconfidence, etc. we tend to invent our own version of the truth [55]. This technique of (sometimes too) critical evaluation has helped us in the past to develop a cautious attitude towards change. Therefore, cognitive psychology recommends to be tolerant for inventions [56,57]. Hence, it was decided to first provide general information for each cognitive variable, including material about the part of the spectrum of the variable the user does not belong to. This general information is limited to a minimum and receives less attention in the tutorial in terms of explanation and rehearsal than the relevant, more specific topics. This way, if a user wants to interpret information differently, (s)he has the background knowledge to do so in a proper context, which should minimize the chance for problematic adaptations.

The resulting training instrument is a digital tutorial that consists of 30 consecutive information screens: a general introduction, 14 screens with information about learning style and the related modeling styles (including exercises and quizzes), 12 screens about field dependency, need for structure and the related guidelines (including exercises and quizzes), and 3 concluding screens with summaries. Fig. 2 shows example screens of the tutorial [54]. More screenshots can be found in the online documentation at <http://www.janclaes.info/papers/SPMM>.

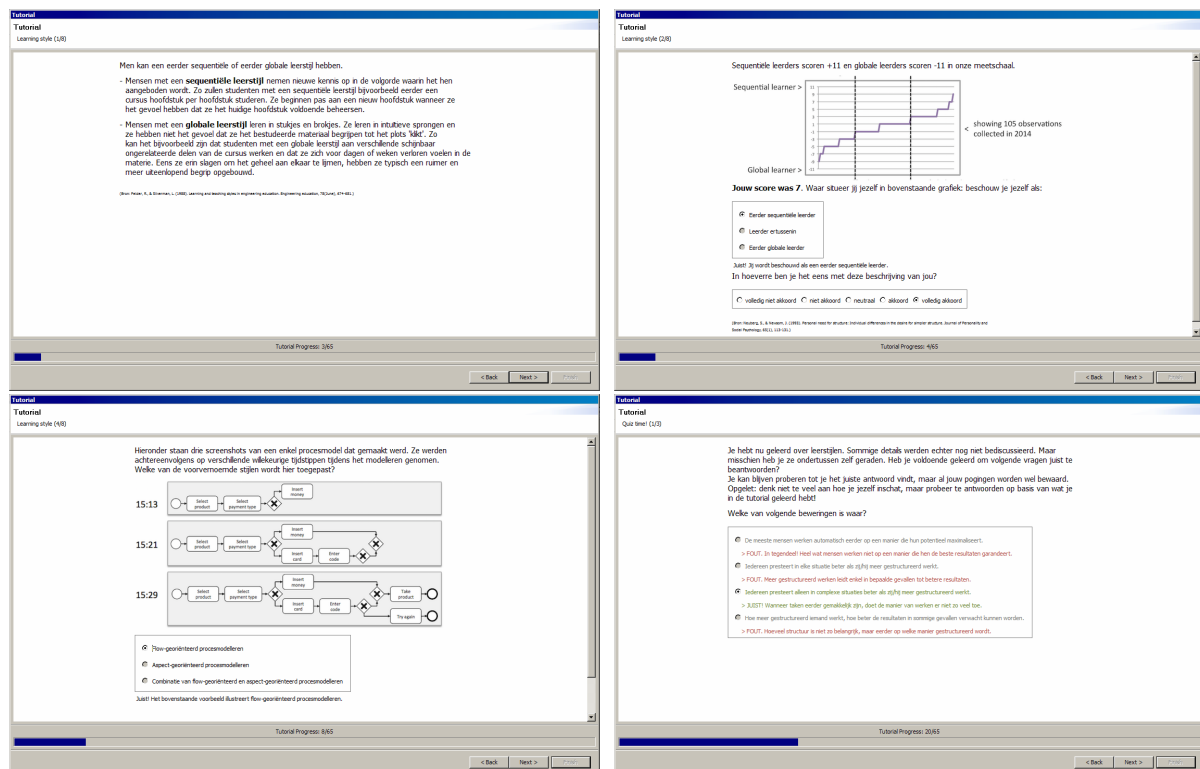


Fig. 2. Example screens from the training instrument (top left: information screen, top right: comprehension question, bottom left: comprehension exercise, bottom right: multiple-choice question showing all feedback)

4 Experimental evaluation of the method and its implementation

In order to evaluate both the method and the supporting implementation, a lab experiment was performed¹. This section describes the tasks (Section 4.1), the participants (Section 4.2), the practical set-up (Section 4.3), the measurements (Section 4.4) and the results (Section 4.5) of the experiment.

¹ At 'http://www.janclaes.info/experiments > 2015GENT' more detailed information about the experiment can be found, as well as zip files containing the constructed models.

4.1 Tasks

The experiment was performed in three parts. In each part of the experiment, the participants had to perform a different set of tasks (see Fig. 3). Except for the Hidden Figures Test and the span tests (see further), there was no task completion limit. After each part of the experiment, a window opened where the user was encouraged to provide feedback on the performed tasks.

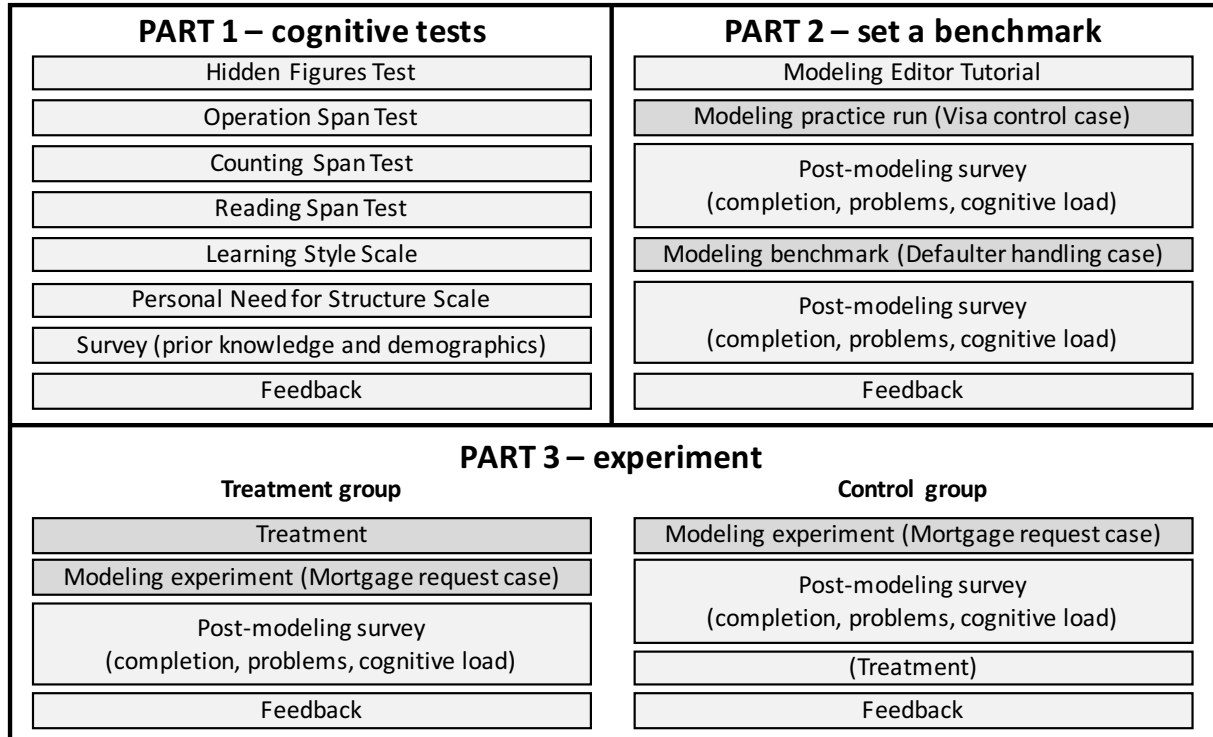


Fig. 3. Overview of the experiment tasks

The first part of the experiment consists of six cognitive tests and a questionnaire about prior knowledge and demographics. The cognitive tests are Hidden Figures Test (field dependency); Operation, Counting and Reading Span tests (working memory capacity as a control variable); Learning Styles scale (learning style); and Personal Need for Structure Scale (desire for structure).

In the second part of the experiment, tasks were used to set a benchmark for every participant in order to be able to compare the modeling results before and after the treatment. The modelers first had to go through the modeling editor tutorial explaining all the features of the simple, BPMN-inspired editor of the experimental environment (see further). Each feature is explained with a description of its use and a short instructional video. In order to assure that the user understood the feature sufficiently, the user had to mimic the example correctly before the tutorial continued. Then, they had to perform

the first modeling task about a visa control process as a practice run, followed by the benchmark modeling task, where the modeler got a case description of a defaulter handling case.

Throughout the experimental flow of tasks, the participants had to create three different process models (i.e., two in the second part and one in the third part). For each of these modeling assignments they were instructed to create a high quality process model based on a textual description of a case, which they received on paper. It was not specified what was meant with a ‘high quality’ model. There was no time limit, but the vast majority of the participants completed each modeling assignment within one hour. Each time, a short post-modeling survey was presented afterwards about task completion, problem occurrences and perceived cognitive load during modeling.

The third part of the experiment consists of a treatment (except for the control group) and the experimental modeling task about a mortgage request process. The treatment task was to go through the training tutorial described in Section 3.2, which automatically selected the appropriate version based on the participant’s cognitive preferences that were measured in the first part of the experiment. For educational reasons, the participants in the control group also retrieved the treatment, but only after all experiment data was collected (see Section 4.3).

4.2 Participants

Master students were selected as ideal study subjects. In contrast to younger students, they are believed to have more mature cognitive reasoning skills and in contrast to practitioners they are a relatively homogeneous group with similar prior knowledge. The former is important for the external validity of the study, whereas the latter relates to the internal validity of comparison between subjects. As prior knowledge is not a variable of the SPMT, experienced process modeling practitioners are not required as study subjects. Two additional benefits of our choice of participants are that their point of cognitive overload may be easier to reach and that they did not suffer from the Expertise Reversal Effect (which was confirmed by their feedback, see further). This is an initial decrease in effectiveness and efficiency when experts are retrained to use another method or tool than they are used to [58]. A more elaborate discussion of the choice to work with master students can be found in [59], p. 141.

In total, 146 master students of the Business Engineering program at Ghent University (Belgium) participated in the experiment. It took place in the context of a master course on Business Process Management. The sessions were planned strategically after the students learned the general principles of process modeling and the syntax of the BPMN language, but before they learned any process modeling method. About two-thirds of the participants are male (62%). The participant's ages range from 21 to 25. About 97% of the participants indicated to have no problems reading and understanding English, although for 99% the mother tongue is Dutch. The answers on dedicated questions confirmed the lack of expertise about process modeling in BPMN and approximated a normal distribution for the questions about the amount of prior knowledge of the modeling domain of the three cases. For 21 students, various kinds of technical or practical problems occurred (e.g., absence, tool and data errors) and they were excluded from the analysis. Excluding these cases, the treatment group consisted of 98 students, whereas the control group contained 27 students. The group allocation was random.

4.3 Practical experiment set-up

The experiment was supported by the Cheetah Experimental Platform (CEP), which is an open source framework for process modeling research [60]. It was selected because of its features of task flow configuration, detailed modeling operation recording and modeling replay functionality. The implementations proposed in this paper (i.e., the cognitive profile measurements and the tutorial, described in Section 3) could easily be integrated in the tool as new kinds of tasks in the experiment flow. Also the middle step of the method (i.e., the selection of a modeling strategy suitable to the cognitive profile) was integrated in the software by automatically selecting the right tutorial variant.

The first part of the experiment, namely the cognitive testing, had to be done at home. The tasks of the first part were introduced in a lecture before the first part took place. In this introduction, the students were instructed to make sure that they would not be distracted (e.g., turn off communication devices and programs, put a 'do not disturb' note on the door). It was assumed that they were intrinsically motivated not to cheat, because of their interest in their own cognitive profile scores. Nevertheless, limited cheat detection was implemented: extremely bad scores were flagged, the time to finish each task was measured and the time the tool was not in focus during the tasks was recorded.

The participants were warned about cheat detection mechanisms, without giving any details though. No other issues were detected than the rare disruptions that were reported in the feedback window. Because of the estimated severity, the results of two participants were thus excluded from the analysis.

The participants had one week to perform the tasks of the first part. They could make each task separately and in any order as long as the answers were submitted in time. Per task, the tool compiled a single zip file with the recorded data. Before closing, it provided clear instructions on how to submit the file correctly to the university's digital learning environment. The whole set-up of the first part was pre-tested in a pseudo-experiment in Eindhoven with 119 students. No technical problems occurred, except for one student claiming to have had submission problems, which could not be verified. Nine students voluntarily participated in post mortem interviews. They confirmed that the instructions were clear and that they recognized themselves in the results of the cognitive tests (see further).

The second and third part of the experiment took place in a computer room of Ghent University. There was one week between these two parts. During this week, it was no problem if the students would talk to each other about the experiment, because the distinction between treatment and control group was only made in part three. For each part, there were two consecutive but further identical sessions to which the students were randomly appointed. Because the whole experimental task flow (including the treatment) was fully automated in the experimental tool, the control group participants could be mixed with the treatment group. As such there was no difference between the experiment circumstances of both groups. In order to minimize distractions, disturbances and influences, both parts and both sessions of each part were closely monitored as if it was an exam session. Finally, in the first lecture after the experiment, the students received collective feedback on the purpose of the experiment and they could download detailed feedback about their individual results from the university's digital learning environment.

Participation to the experiment was voluntary. By participating the student agreed that data about their cognitive profile and modeling behavior would be collected, together with their answers to the questions of the surveys. The students were motivated to participate by receiving a bonus point for the

BPM course. Further, they were also stimulated to perform the best possible, because all participants had a different chance to win an iPhone, depending on how well they followed the instructions. Each student started with five chances, but could lose a chance when a quality issue was detected. The five quality issues that were monitored are: missing or late submissions, cheat detection, poor answers on the open questions, little effort to adapt modeling strategy as instructed, other problems. The choice to spend the budget to a single grand prize instead of multiple smaller prizes was made after questioning a dozen students not enrolled in the course about what would motivate them more. Further, the participants were told they could stop at any point in time. Students that were not able to participate could earn the bonus point with a replacement task. Furthermore, the students of the control group got the treatment at the end of the experiment. This way, they also learned about their optimal process modeling strategy, but only after their experimental data was collected.

4.4 Measurements

For several variables a measure needed to be selected from different alternatives, to be adapted to our needs or to be invented, because a standardized and validated measure does not exist. Except for the described measures for the three cognitive variables used in the method (see Section 3.1), other variables were calculated to evaluate the method and its implementation. Their measurement is described in an appendix, because it is not the goal of this paper to propose new metrics (see Appendix D). Construct validity is not warranted because the validity and reliability of the measurements and derived metrics are not known. Also, no effort was made to further study or improve their validity. Nevertheless, it is assumed that imperfect measures rather have a negative impact on the size and significance of the tested relations. Therefore, the results described hereafter may be too negative.

4.5 Results

This subsection discusses the analysis techniques and results for the evaluation of the treatment adoption (Section 4.5.1), the treatment effect (Section 4.5.2) and the user perceptions (Section 4.5.3).

4.5.1 Treatment adoption

First, the treatment adoption was investigated. Did the participants in the treatment group change their behavior as instructed by their individualized tutorial? In order to determine this, the results of the benchmark case were compared to the results of the experimental case with the cognitive profile of the modeler in mind and these effects were then compared between treatment and control group (i.e., combination of within-subject and between-subjects comparison). The results are shown in Fig. 4.

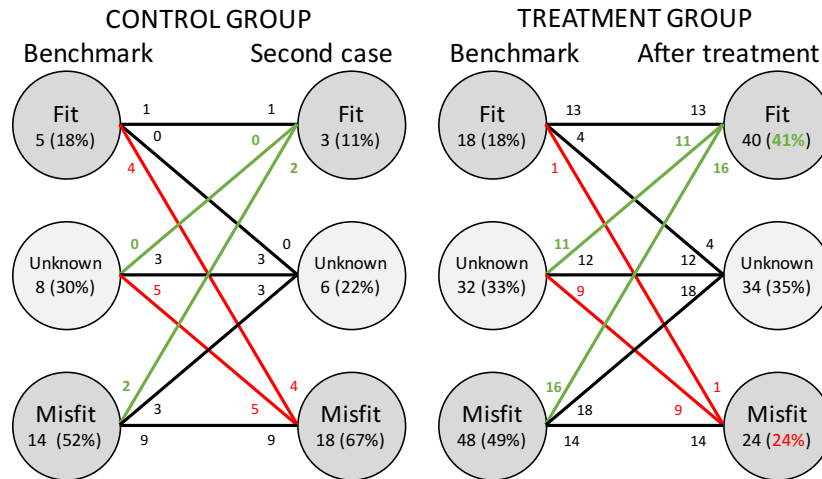


Fig. 4. Results for treatment adoption

The following observations can be made.

- In the control group it can be observed how some participants remained in their category (see numbers near horizontal lines), whereas others changed their sequencing approach. Perhaps these participants may have preferred another approach depending on the case.
- In the control group, regardless the various changes of category, it can also be observed how the distribution over categories is relatively constant.
- In the treatment group however, there was a bigger difference: the number of fitting techniques increased from 18% to 41%, whereas the misfitting techniques decreased from 49% to 24%.
- Not every participant improved after the treatment. Some participants remained in their category (see numbers near horizontal lines) or even moved to a worse one (see numbers near red lines).
- Because we do not know for sure how well the participants in the 'Unknown' category performed, it is interesting to see how the results change when they are left out of the analysis. In the control group the number of fitting approaches then decreases from 5/16 (31%) to 3/16 (19%),

whereas the misfitting approaches increase from 11/16 (69%) to 13/16 (81%). In contrast, in the treatment group the fitting approaches increase from 14/44 (32%) to 29/44 (65%), whereas the misfitting approaches decrease from 30/44 (68%) to 15/44 (34%).

Because of the relatively low size of participants in each group and category, it is difficult to use these numbers for generalization, but it can still be concluded that they indicate a positive treatment adoption. The method appears to be successful in directing the participants towards the optimal sequencing technique.

4.5.2 Treatment effect

In the previous subsection it was established that the treatment increases the fit of the applied sequencing technique, at least for a considerable part of the participants. According to the SPMT this should have a positive effect on cognitive (over)load and thus also on process modeling effectiveness (i.e., number of errors) and efficiency (i.e., time and effort). The degree of sequencing is expected to have both a positive and a negative effect on the cognitive load [40] and thus it depends on which of both effects is bigger to determine the net effect. Next, the structuredness of sequencing and the fit of the overall modeling strategy are expected to have only beneficial effects on cognitive load (more structuredness and more fit leading to less cognitive load) [40]. Therefore, in this subsection the focus is on the relation between the applied overall process modeling strategy and the cognitive load and resulting modeling efficacy. The presented numbers below are from the experimental case, but similar results were obtained for the benchmark case.

The results of the statistical study on the relations with cognitive load were inconclusive. No significant relation was found between the degree, structuredness and fit of modeling and the overall cognitive load. Moreover, the more fine-grained dual-task metric for cognitive load could not be used, because of a too small variation of the values. Because the one metric for cognitive load is too general (but validated) and the other metric is not validated (but more precise), these problems were anticipated and next to cognitive load four additional variables were used that measure modeling effectiveness and efficiency (recall that cognitive overload causes an increase in mistakes and a

decrease in speed, see Section 2.2). Therefore, the syntactic and semantic quality of the produced process model were calculated, as well as the overall modeling time and effort (see Appendix D for an overview of the calculated metrics). The statistical analyses with these four dependent variables revealed interesting results. They are discussed below.

A. Bivariate correlations

First, in order to investigate the relations between each pair of variables separately, simple bivariate correlations were calculated. The independent variables are degree, structuredness and fit and the dependent variables are syntactic errors, syntactic error types, semantic errors, semantic error types, time and effort. The results are presented in Table 1.

Table 1. Bivariate correlations: two-tailed spearman's rho

N=122	syntactic errors	syntactic error types	semantic errors	semantic error types	modeling time	modeling effort
degree	-0,073 (p=0,424)	-0,074 (p=0,419)	** -0,239 (p=0,008)	* -0,225 (p=0,013)	** 0,295 (p=0,001)	** 0,710 (p=0,000)
structuredness	-0,174 (p=0,055)	-0,164 (p=0,071)	0,175 (p=0,053)	0,018 (p=0,842)	** -0,363 (p=0,000)	** -0,448 (p=0,000)
fit	-0,124 (p=0,174)	-0,111 (p=0,222)	-0,083 (p=0,366)	-0,109 (p=0,232)	-0,024 (p=0,794)	* -0,226 (p=0,012)

The following observations can be made.

- The data indicates that the net effect of the degree of sequencing on modeling quality is a negative, which is beneficial for the modeler (more sequencing means less errors). Further, an increasing degree of sequencing relates to an increasing modeling time, which also makes sense (the more pauses, the more time). More surprising is the positive effect on effort. We find it hard to explain this observation, although it does not contradict with the SPMT. Perhaps the direction of causality is reversed: modelers that used more operations for modeling (=‘effort’), take more time and have a higher number of pauses (=‘degree’) under the definitions of our metrics?
- For structuredness there were only significant relations with time and effort. The more structured, the less time and effort. The effect on modeling quality is mixed: a negative effect on syntactic errors and a positive on semantic errors. These results are not significant though.
- A similar observation can be made for fit, albeit that only the relation with effort is significant. Note how the signs are in the expected direction: the more fit, the less errors, time and effort.

- The lower in Table 1, the less significant the results. This is because the lower metrics are less accurate and the lower independent variables are included in the higher ones.

Summarized, all the significant findings support the claim, but a concern can be raised about the relation between degree and effort which accords to the SPMT theory, but to us seems contra-intuitive.

B. Stepwise linear regression

Next, the aggregated effect of the three independent variables was investigated with a stepwise linear regression. Such a regression can reveal if a combination of the three modeling strategy factors could be used to model the variation in the dependent variables. The technique warrants significant results (p-values less than 0,05). The results are presented in Table 2. At the bottom, the R^2 value and the p-value of the complete model are given. The R^2 value indicates how much of the variation in the dependent variable can be explained with the variation in the independent variables, according to the presented model. The table can be read as follows (consider the last column): 56,60% of the variation in modeling effort can be explained by the variation in the degree and structuredness of the applied technique with the formula $effort = 143,155 + (190,559 \times Degree) + (-8,553 \times Structuredness)$.

Table 2. Stepwise linear regression with three independent variables: unstandardized beta coefficients

	syntactic errors	syntactic error types	semantic errors	semantic error types	modeling time	modeling effort
constant	2,875 (p=0,000)	1,511 (p=0,000)	7,369 (p=0,000)	3,321 (p=0,000)	42,636 (p=0,000)	143,155 (p=0,000)
degree	-	-	-2,908 (p=0,010)	-1,119 (p=0,034)	10,067 (p=0,014)	190,559 (p=0,000)
structuredness	-0,508 (p=0,002)	-0,2 (p=0,016)	-	-	-2,023 (p=0,003)	-8,553 (p=0,005)
fit	-	-	-2,182 (p=0,038)	-	-	-
R^2	7,90% (p=0,002)	4,70% (p=0,016)	7,60% (p=0,009)	3,70% (p=0,034)	15,20% (p=0,000)	56,60% (p=0,000)

The following observations can be made.

- All the signs are as expected. The higher the degree, structuredness and fit, the less errors are made. The higher the degree of sequencing, the more time and effort it takes. The higher the structuredness of sequencing, the less time and effort it takes. Note how the effect between degree and effort is again positive (in combination with the other variables in the regression model).

- A varying subset of only these three independent variables (i.e., degree, structuredness and fit) account for 7,5% to 8% of the variation in errors, about 15% of the variation in time and more than 55% of the variation in effort.

These are considered good results. The outcomes support the SPMT and they also showcase the usefulness of the method by demonstrating the effect of applying an optimal process modeling strategy on the modeling effectiveness and efficiency.

C. Extended set of independent variables

Further, with the aim to improve the explanatory power of the regression models, the set of independent variables was extended with the cognitive preference variables (i.e., learning style, field dependency and need for structure) to examine if they could also have a direct effect on modeling efficacy. Moreover, also the dependent variables themselves were considered as independent for the other dependent variables in order to reveal interaction effects. The results are presented in Table 3.

Table 3. Stepwise linear regression with multiple independent variables: unstandardized beta coefficients

	syntactic errors	syntactic error types	semantic errors	semantic error types	modeling time	modeling effort
constant	2,033 (p=0,000)	1,126 (p=0,000)	9,072 (p=0,000)	3,597 (p=0,000)	22,872 (p=0,000)	163,511 (p=0,000)
degree	-	-	-	-	-10,43 (p=0,040)	177,224 (p=0,000)
structuredness	-0,518 (p=0,001)	-0,205 (p=0,012)	-	-	-	-9,027 (p=0,003)
field dependency	2,274 (p=0,004)	1,04 (p=0,011)	-	0,978 (p=0,026)	-	-32,809 (p=0,030)
time	-	-	-0,099 (p=0,000)	-0,028 (p=0,013)	-	-
effort	-	-	-	-	0,117 (p=0,000)	-
R ²	14,10% (p=0,000)	9,70% (p=0,002)	12,80% (p=0,000)	10,40% (p=0,001)	32,70% (p=0,000)	58,40% (p=0,000)

The following observations can be made.

- All the models got more accurate (i.e., increased R² values).
- About 13% of the variation in semantic errors can be related to only the variation in time. This is interesting. Apparently, the more time was utilized, the less semantic errors were made.
- Also, the variation in time can for about 33% be explained by the variation in degree of sequencing and effort.

Combining these latter two observations, it may be the case that more effort leads to more time, which in turn leads to less errors. The idea arises that some modelers may have used the extra time and effort to correct semantic mistakes. There may be a tradeoff between time/effort and (the semantic aspect of) quality. This is a sensible hypothesis because it can be explained as follows. When the modeler is overloaded, mistakes are made. Later, when the cognitive load drops to manageable proportions, the overload disappears and the modeler can use the freed capacity to correct the mistakes. Therefore, this hypothesis is interesting and will be tested in future research.

4.5.3 Perceived usefulness and ease of use

Besides the evaluation of treatment adoption (did modelers change their behavior after the limited intervention) and the treatment effect (did applying a more fitting process modeling technique decrease the number of modeling mistakes, time and effort), we are also interested in the user perceptions to get a preliminary idea about the actual use of the method and implementation. Two sources were used to assess the perceived usefulness and ease of use: the feedback that was collected in the tool and answers to interview questions. The applied interviewing protocol is presented in Appendix H. This way, an extensive set of 495 feedback comments were collected. The conclusions are presented below. They are illustrated with concrete quotes in Appendix I.

A. Perceptions about the automation of the cognitive tests

First, the feedback on the cognitive tests was examined. Most participants seem to have enjoyed to perform the cognitive tests. There were no negative *motivational* comments other than that the tests were hard and required a lot of concentration. The vast majority of *technical* comments indicated that there were no interruptions or technical problems. Most of the rare disruptions seem to have been minimal and were quickly resolved. However, for some students it was not clear how much the disturbance has influenced the results and they were left out from further analysis. The *content* of the implementation was perceived as clear, especially after the demo exercise. All in all, it is concluded that the automation of the cognitive tests and the unsupervised setting was perceived as fairly good.

B. Perceptions about the training instrument

There were no surprising comments about the technical aspects of the training instrument. When a video was streamed over the internet or when the tool was sending much information to our database, the tool was rather slow. The students were warned in advance for this and there were no problematic reactions about this issue. In contrast, a high number of positive and surprised reactions were received. A number of participants recognized aspects explained by the tutorial, whereas others discovered new insights about their own personality. Although some participants were surprised about what they learned, nobody objected to the provided knowledge and guidelines in their feedback. Only one comment was received about some unclear explanation. No other issues were mentioned or detected.

C. Perceptions about the results of the method

The majority of the participants were optimistic about the overall method. They indicated to have the impression it has helped them to improve their modeling. At the negative side, a number of participants complained about the complexity of the assignment, especially towards the end. This is normal because cognitive load increases during modeling. In fact, they were probably experiencing cognitive overload. This would mean that the method did not help to implement a strategy that avoids overload completely (which was also never assumed). There were two comments about already applying the proposed modeling technique. Indeed, some people automatically apply the best suitable strategy and will not experience any improvement with the developed method.

5 Discussion

The research discussed in this paper can potentially have a direct impact on process modeling practice. Although further investigation is required of different aspects of the developed method, the results are promising and we see no reason why the method cannot be already applied in real-world application situations. As such, it can help novice modelers to become aware of their strengths and weaknesses as a modeler and to learn a modeling strategy that is good for them. Unless there is reason to believe that the results presented here apply only to modeling novices, the method can also be applied by experienced practitioners. This way, the practical value of the method is that it can improve

both the training and the application of process modeling. This is relevant as there is clearly a need for more smart business process management and modeling tools (see Section 2.1).

There is also a potential impact on research. The automation of the psychological measurements that we propose in this paper may simplify various studies. The contribution is both technical and conceptual. The program code (i.e. technical contribution) is accompanied by directions about motivation, cheat detection, disturbance avoidance and lack of supervision (i.e., conceptual contribution). Further, the research demonstrates how a sophisticated training instrument can successfully replace certain more extensive and thus more labor-intensive treatments, which is also beneficial for the reproducibility, the objectiveness of the description and the spread of the research.

Nevertheless, there are a number of limitations to the research in its current form. Because of the extensive size of the project, it was not feasible and out of scope to perfect every detail of the research.

First, validated measurements are lacking for the concepts of sequencing degree, sequencing structuredness, overall modeling strategy fit, cognitive overload, instantaneous cognitive load, cognitive syntactic errors and cognitive semantic errors. An attempt was made to use the best available measurements and adapt them as little as possible in order to be confident in the results. However, the construct validity of these metrics can currently not be guaranteed. Nevertheless, the results obtained with these metrics were critically evaluated and still provide promising indications that should not be dedicated to pure coincidence.

Second, for the reasons mentioned in Section 4.2 and 4.3 the research was performed as a lab experiment with an artificial set-up. In other words, the ecological validity is not known. In a training context, the participants may receive more guidance of a supporting lecturer and will experience a lower mental burn due to the other experimental tasks when utilizing the method or training instrument for training a modeling strategy. In a modeling environment, the user is a more experienced practitioner and additional effects such as tool habituation, modeling conventions, managerial or operational influences can distort the results. A more realistic environment and conditions should be implemented in order to study the effects in their proper context.

6 Conclusion

As mentioned in the introduction, more and more process models are constructed in organizations. At the same time, the quality of process models is low in many cases. Therefore, a smart method was developed to assist modelers to discover and train a process modeling strategy that helps them to construct a high-quality process model. The focus is not on lacking or imperfect prior knowledge about the process or about modeling, which can be avoided with targeted training in order to complete the missing knowledge. On the contrary, the focus is on issues with the cognitive processing of available knowledge. By suggesting and supporting the implementation of an optimal approach, the method helps the modeler in her/his cognitive task of translating the mental image about the process in a syntactic and semantic correct formal process representation (i.e., the process model).

The developed method consists of three steps: (i) measure the cognitive profile of the modeler, (ii) determine the best fitting process modeling strategy, and (iii) train this strategy. The three steps are supported in a digital implementation that can be run autonomously by the user. To achieve this, an automated version of existing cognitive measurements was developed, as well as a digital, differentiated and adaptive training instrument.

The artefacts (i.e., the conceptual method and its supporting implementation) were jointly evaluated with a large-scale laboratory experiment. The experiment was pretested with 119 master students in Eindhoven and ran in the context of a Business Process Management course with 149 master student participants in Ghent. Before, during and after the experiment, various variables were measured quantitatively and/or qualitatively. The results are merely positive. They indicate that (i) the training instrument succeeded well in changing the participant's modeling behavior as intended with a rather limited intervention, (ii) the application of the method helps improving the efficiency of modeling, (iii) the method did not directly improve the effectiveness of modeling, but a trade-off between efficiency and effectiveness is hypothesized, (iv) the participants perceived the method and its tool support as useful.

In future research, we will investigate if and how the method can be made even more smart. For example, it can be interesting to monitor the modeler while modeling and to occasionally interrupt with feedback and adjustment messages. A number of the measures that are currently used only to evaluate the method will than probably become part of the method. Or, the research can be connected to existing work on ontological process modeling support (see for example the overview of [61] or our work on smart ontology-supported conceptual modeling [62]).

Future research will also target the limitations discussed in Section 5. We will further investigate how a distinction between knowledge and cognition related quality issues can be made. As possible approach, during an individual feedback session the modeler and researcher can run through the modeling replay and together they can determine the cause for each error. Also, because it cannot be easily measured when cognitive overload occurs, the measures for cognitive load need to be further explored and perfected. Having an instantaneous measure of cognitive load will allow for a more fine-grained analysis. These metrics can also support the investigation of the formulated hypothesis that there is a trade-off between modeling efficiency and modeling effectiveness.

Next, it is currently not clear if and to what extent the results of this research apply in various other settings. Therefore, a selection of different cases will be examined further. We will investigate how the method performs for modelers that operate in a context of other languages (including the full BPMN modeling language), that use other modeling tools, and for other users (e.g., modeling practitioners), other modeling cases, etc. Then, the scope of the research will be further expanded by including process model understanding by model readers or by people handling various other similar problem solving assignments (such as conceptual modeling, programming or text writing).

Acknowledgements

We thank the many experiment participants, as well as the 7 administrators, the 13 raters and the 9 interviewees both in Eindhoven and in Ghent for their valuable contribution to the research.

References

- [1] I. Davies, P. Green, M. Rosemann, M. Indulska, S. Gallo, How do practitioners use conceptual modeling in practice?, *Data Knowl. Eng.* 58 (2006) 358–380. doi:10.1016/j.datak.2005.07.007.
- [2] M. Indulska, J.C. Recker, M. Rosemann, P. Green, Business process modeling: Current issues and future challenges, in: P. Van Eck, J. Gordijn, R. Wieringa (Eds.), *Proc. 21st Int. Conf. Adv. Inf. Syst. Eng. (CAiSE '09)*, Amsterdam, Netherlands, June 8-12, 2009, Springer, 2009: pp. 501–514. doi:10.1007/978-3-642-02144-2_39.
- [3] N. Hassan, J.C. Recker, E. Bernhard, *A study of the use of business process modelling at suncorp*, Brisbane, Australia, 2011.
- [4] V. Gruhn, R. Laue, What business process modelers can learn from programmers, *Sci. Comput. Program.* 65 (2007) 4–13. doi:10.1016/j.scico.2006.08.003.
- [5] J. Mendling, H.M.W. Verbeek, B.F. Van Dongen, W.M.P. Van der Aalst, G. Neumann, Detection and prediction of errors in EPCs of the SAP reference model, *Data Knowl. Eng.* 64 (2008) 312–329. doi:10.1016/j.datak.2007.06.019.
- [6] O.I. Lindland, G. Sindre, A. Solvberg, Understanding quality in conceptual modeling, *IEEE Softw.* 11 (1994) 42–49. doi:10.1109/52.268955.
- [7] H.J. Nelson, G. Poels, M. Genero, M. Piattini, A conceptual modeling quality framework, *Softw. Qual. J.* 20 (2012) 201–228. doi:10.1007/s11219-011-9136-9.
- [8] S. Rockwell, A. Bajaj, COGEVAL: Applying cognitive theories to evaluate conceptual models, *Adv. Top. Database Res.* 4 (2005) 255–282. doi:10.4018/978-1-59140-471-2.ch012.
- [9] J. Krogstie, G. Sindre, H. Jørgensen, Process models representing knowledge for action: A revised quality framework, *Eur. J. Inf. Syst.* 15 (2006) 91–102. doi:10.1057/palgrave.ejis.3000598.
- [10] J. Becker, M. Rosemann, C. Von Uthmann, Guidelines of business process modeling, in: W.M.P. Van der Aalst, J. Desel, A. Oberweis (Eds.), *Bus. Process Manag. Model. Tech. Empir. Stud. Part I.*, Springer-Verlag Berlin Heidelberg, 2000: pp. 30–49. doi:10.1007/3-540-45594-9_3.

- [11] R. Laue, J. Mendling, Structuredness and its significance for correctness of process models, *Inf. Syst. E-Bus. Manag.* 8 (2010) 287–307. doi:10.1007/s10257-009-0120-x.
- [12] H.A. Reijers, J. Mendling, A study into the factors that influence the understandability of business process models, *IEEE Trans. Syst. Man, Cybern. Part A Syst. Humans.* 41 (2011) 449–462. doi:10.1109/TSMCA.2010.2087017.
- [13] V. Gruhn, R. Laue, Adopting the cognitive complexity measure for business process models, in: Y.Y. Yao, Z.Z. Shi, Y. Wang, W. Kinsner (Eds.), 5th IEEE Int. Conf. Cogn. Informatics, IEEE, 2006: pp. 236–241. doi:10.1109/COGINF.2006.365702.
- [14] L. Sánchez-González, F. García, F. Ruiz, J. Mendling, Quality indicators for business process models from a gateway complexity perspective, *Inf. Softw. Technol.* 54 (2012) 1159–1174. doi:10.1016/j.infsof.2012.05.001.
- [15] J.C. Recker, J. Mendling, Adequacy in process modeling: A review of measures and a proposed research agenda-position paper, in: CAiSE 2007 Work. Proc. Vol. 1, 2007: pp. 235–244.
- [16] I. Vanderfeesten, J. Cardoso, J. Mendling, H.A. Reijers, W.M.P. Van der Aalst, Quality metrics for business process models, in: L. Fisher (Ed.), *BPM Work. Handb., Future Strategies*, 2007: pp. 179–190.
- [17] M. Van Mersbergen, A framework for business process model quality and an evaluation of model characteristics as predictors for quality, Technische Universiteit Eindhoven, 2013.
- [18] J. Mendling, Detection and prediction of errors in EPC business process models, Vienna University of Economics and Business Administration, 2007.
- [19] J.C. Recker, A. Dreiling, Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education, in: *ACIS 2007 Proc.*, 2007: pp. 356–366.
- [20] H.A. Reijers, J. Mendling, J.C. Recker, On the usage of labels and icons in business process modeling, in: J. Vom Brocke, M. Rosemann (Eds.), *Handb. Bus. Process Manag.* 1, Springer-Verlag Berlin Heidelberg, 2010: pp. 167–185. doi:10.4018/jismd.2010040103.
- [21] I. Moreno-Montes de Oca, M. Snoeck, H.A. Reijers, A. Rodríguez-Morffi, A systematic

- literature review of studies on business process modeling quality, *Inf. Softw. Technol.* 58 (2014) 187–205. doi:10.1016/j.infsof.2014.07.011.
- [22] J. Mendling, H.A. Reijers, W.M.P. Van der Aalst, Seven process modeling guidelines (7PMG), *Inf. Softw. Technol.* 52 (2010) 127–136. doi:10.1016/j.infsof.2009.08.004.
- [23] J. Mendling, L. Sánchez-González, F. García, M. La Rosa, Thresholds for error probability measures of business process models, *J. Syst. Softw.* 85 (2012) 1188–1197. doi:10.1016/j.jss.2012.01.017.
- [24] M. La Rosa, P. Wohed, J. Mendling, A.H.M. Ter Hofstede, H.A. Reijers, W.M.P. Van der Aalst, Managing process model complexity via abstract syntax modifications, *IEEE Trans. Ind. Informatics.* 7 (2011) 614–629. doi:10.1109/TII.2011.2166795.
- [25] M. La Rosa, A.H.M. Ter Hofstede, P. Wohed, H.A. Reijers, J. Mendling, W.M.P. Van der Aalst, Managing process model complexity via concrete syntax modifications, *IEEE Trans. Ind. Informatics.* 7 (2011) 255–265. doi:10.1109/TII.2011.2124467.
- [26] B. Silver, *BPMN: Method & style*, 2nd ed., Cody-Cassidy Press, 2011.
- [27] B. Silver, *BPMS Watch: Ten tips for effective process modeling*, (n.d.).
<http://www.bpminstitute.org/resources/articles/bpms-watch-ten-tips-effective-process-modeling>.
- [28] K. Worden, W.A. Bullough, J. Haywood, *Smart technologies*, World Scientific, 2003. doi:10.1142/4832.
- [29] Y. Wand, R. Weber, An ontological model of an information system, *IEEE Trans. Softw. Eng.* 16 (1990) 1282–1292. doi:10.1109/32.60316.
- [30] J.C. Recker, A socio-pragmatic constructionist framework for understanding quality in process modelling, *Australas. J. Inf. Syst.* 14 (2007) 43–63. doi:10.3127/ajis.v14i2.23.
- [31] W. Bandara, G.G. Gable, M. Rosemann, Factors and measures of business process modelling: model building through a multiple case study, *Eur. J. Inf. Syst.* 14 (2005) 347–360. doi:10.1057/palgrave.ejis.3000546.
- [32] M. Vandewaetere, P. Desmet, G. Clarebout, The contribution of learner characteristics in the development of computer-based adaptive learning environments, *Comput. Human Behav.* 27

- (2011) 118–130. doi:10.1016/j.chb.2010.07.038.
- [33] P. Subban, Differentiated instruction: A research basis, *Int. Educ. J.* 7 (2006) 935–947. doi:10.1111/j.1365-2648.2006.04074.x.
- [34] P. Brusilovsky, Adaptive hypermedia, *User Model. User-Adapt. Interact.* 11 (2001) 87–110. doi:10.1023/A:1011143116306.
- [35] P. Karampiperis, D. Sampson, Adaptive learning resources sequencing in educational hypermedia systems, *Educ. Technol. Soc.* 8 (2005) 128–147.
- [36] J.C.R. Tseng, H.C. Chu, G.J. Hwang, C.C. Tsai, Development of an adaptive learning system with two sources of personalization information, *Comput. Educ.* 51 (2008) 776–786. doi:10.1016/j.compedu.2007.08.002.
- [37] N. Ford, Cognitive styles and virtual environments, *J. Am. Soc. Inf. Sci.* 51 (2000) 543–557. doi:10.1002/(SICI)1097-4571(2000)51:6<543::AID-ASI6>3.0.CO;2-S.
- [38] E. Triantafillou, A. Pomportsis, S. Demetriadis, The value of adaptivity based on cognitive style : an empirical study, *Br. J. Educ. Psychol.* 35 (2004) 95–106. doi:10.1111/j.1467-8535.2004.00371.x.
- [39] J. Claes, I. Vanderfeesten, H.A. Reijers, J. Pinggera, M. Weidlich, S. Zugal, D. Fahland, B. Weber, J. Mendling, G. Poels, Tying process model quality to the modeling process: The impact of structuring, movement, and speed, in: A. Barros, A. Gal, E. Kindler (Eds.), *Proc. BPM'12, LNCS 7481*, Springer, 2012: pp. 33–48.
- [40] J. Claes, I. Vanderfeesten, F. Gailly, P. Grefen, G. Poels, The Structured Process Modeling Theory (SPMT) - A cognitive view on why and how modelers benefit from structuring the process of process modeling, *Inf. Syst. Front.* 17 (2015) 1401–1425. doi:10.1007/s10796-015-9585-y.
- [41] J. Claes, I. Vanderfeesten, J. Pinggera, H.A. Reijers, B. Weber, G. Poels, A visual analysis of the process of process modeling, *Inf. Syst. E-Bus. Manag.* 13 (2015) 147–190. doi:10.1007/s10257-014-0245-4.
- [42] J. Sweller, Cognitive load during problem solving: Effects on learning, *Cogn. Sci.* 12 (1988) 257–285. doi:10.1016/0364-0213(88)90023-7.

- [43] P. Gerjets, K. Scheiter, R. Catrambone, Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures, *Instr. Sci.* 32 (2004) 33–58. doi:10.1023/B:TRUC.0000021809.10236.71.
- [44] R.M. Felder, L.K. Silverman, Learning and teaching styles in engineering education, *Eng. Educ.* 78 (1988) 674–681.
- [45] H.A. Witkin, C.A. Moore, D.R. Goodenough, P.W. Cox, Field-dependent and field-independent cognitive styles and their educational implications, *Rev. Educ. Res.* 47 (1977) 1–64. doi:10.1002/j.2333-8504.1975.tb01065.x.
- [46] S.L. Neuberg, J.T. Newsom, Personal need for structure: Individual differences in the desire for simpler structure., *J. Pers. Soc. Psychol.* 65 (1993) 113–131. doi:10.1037/0022-3514.65.1.113.
- [47] J. Claes, I. Vanderfeesten, F. Gailly, P. Grefen, G. Poels, Towards a structured process modeling method: Building the prescriptive modeling theory (accepted), in: *Proc. BPM 2016 Conf. Work.*, Springer, 2017.
- [48] S. Gregor, The nature of theory in information systems, *MIS Q.* 30 (2006) 611–642.
- [49] G. Pask, Learning strategies, teaching strategies, and conceptual or learning style, in: R.R. Schmeck (Ed.), *Learn. Strateg. Learn. Styles*, Springer, 1988: pp. 83–100. doi:10.1007/978-1-4899-2118-5_4.
- [50] N. Ford, Learning styles and strategies of postgraduate students, *Br. J. Educ. Technol.* 16 (1985) 65–77. doi:10.1111/j.1467-8535.1985.tb00483.x.
- [51] H.A. Witkin, *A manual for the Embedded Figures Test*, Consulting Psychologists Press, Palo Alto, CA, 1971.
- [52] R.B. Ekstrom, J.W. French, H.H. Harman, *Kit of factor-referenced cognitive tests*, Princeton, New Jersey, 1976.
- [53] M.M. Thompson, M.E. Naccarato, K.C.H. Parker, G.B. Moskowitz, The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions, in: G. Moskowitz (Ed.), *Cogn. Soc. Psychol. Princet. Symp. Leg. Futur. Soc. Cogn.*, Erlbaum, Mahwah, NJ, 2001: pp. 19–39.
- [54] J. Claes, *Structured Process Modeling Method experimental tool*, (2016).

doi:10.5281/zenodo.166727.

- [55] J. Klayman, Varieties of Confirmation bias, *Psychol. Learn. Motiv. - Adv. Res. Theory*. 32 (1995) 385–418. doi:10.1016/S0079-7421(08)60315-1.
- [56] M.F. Young, Instructional design for situated learning, *Educ. Technol. Res. Dev.* 41 (1993) 43–58. doi:10.1007/BF02297091.
- [57] D.H. Jonassen, Instructional design models for well-structured and ill-structured problem-solving learning outcomes, *Educ. Technol. Res. Dev.* 45 (1997) 65–90. doi:10.1007/BF02299613.
- [58] S. Kalyuga, P. Ayres, P. Chandler, J. Sweller, The expertise reversal effect, *Educ. Psychol.* 38 (2003) 23–31. doi:10.1207/S15326985EP3801_4.
- [59] J. Claes, Investigating the process of process modeling and its relation to modeling quality - The role of structured serialization, Ghent University, Eindhoven University of Technology, 2015.
- [60] J. Pinggera, S. Zugall, B. Weber, Investigating the process of process modeling with Cheetah Experimental Platform, in: B. Mutschler, J. Recker, R. Wieringa, J. Ralyte, P. Plebani (Eds.), *Proc. 1st Int. Work. Empir. Res. Process. Inf. Syst. (ER-POIS '10)*, Hammamet, Tunis. June 8, 2010, CEUR-WS, 2010: pp. 13–18.
- [61] J.B. Gassen, J. Mendling, A. Bouzeghoub, L.H. Thom, J.P.M. De Oliveira, An experiment on an ontology-based support approach for process modeling, *Inf. Softw. Technol.* 0 (2016) 1–22. doi:10.1016/j.infsof.2016.11.005.
- [62] F. Gailly, N. Alkhaldi, S. Casteleyn, W. Verbeke, Recommendation-based Conceptual Modelling and Ontology Evolution Framework (CMOE+), *Bus. Inf. Syst. Eng.* (in press) (2016) 1–36.
- [63] F.G.W.C. Paas, J.E. Tuovinen, H. Tabbers, P.W.M. Van Gerven, Cognitive load measurement as a means to advance cognitive load theory, *Educ. Psychol.* 38 (2003) 63–71. doi:10.1207/S15326985EP3801_8.
- [64] M. Lansman, E. Hunt, Individual differences in secondary task performance, *Mem. Cognit.* 10 (1982) 10–24. doi:10.3758/BF03197621.

Appendix A. Learning style – Questionnaire

The following questionnaire is used to make a distinction between sequential and global learners. It is adapted from <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>: 33 questions were left out, because they are not related to the sequential/global learner distinction and this significantly decreases the mental effort for the respondent to fill out the questionnaire. It is assumed that this has a negligible effect on the validity of the measurement scale. The resulting score is an odd value between -11 (global learner) and +11 (sequential learner) and is calculated as the number of times the respondent answered (a) minus the number of answers (b).

1. I tend to
 - (a) understand details of a subject but may be fuzzy about its overall structure.
 - (b) understand the overall structure but may be fuzzy about details.
2. Once I understand
 - (a) all the parts, I understand the whole thing.
 - (b) the whole thing, I see how the parts fit.
3. When I solve math problems
 - (a) I usually work my way to the solutions one step at a time.
 - (b) I often just see the solutions but then have to struggle to figure out the steps to get to them.
4. When I'm analyzing a story or a novel
 - (a) I think of the incidents and try to put them together to figure out the themes.
 - (b) I just know what the themes are when I finish reading and then I have to go back and find the incidents that demonstrate them.
5. When I start a homework problem, I am more likely to
 - (a) start working on the solution immediately.
 - (b) try to fully understand the problem first.
6. It is more important to me that an instructor
 - (a) lay out the material in clear sequential steps.
 - (b) give me an overall picture and relate the material to other subjects.
7. I learn
 - (a) at a fairly regular pace. If I study hard, I'll "get it."
 - (b) in fits and starts. I'll be totally confused and then suddenly it all "clicks."
8. When considering a body of information, I am more likely to
 - (a) focus on details and miss the big picture.
 - (b) try to understand the big picture before getting into the details.
9. When writing a paper, I am more likely to
 - (a) work on (think about or write) the beginning of the paper and progress forward.
 - (b) work on (think about or write) different parts of the paper and then order them.
10. Some teachers start their lectures with an outline of what they will cover. Such outlines are
 - (a) somewhat helpful to me.
 - (b) very helpful to me.
11. When solving problems in a group, I would be more likely to
 - (a) think of the steps in the solution process.
 - (b) think of possible consequences or applications of the solution in a wide range of areas.

Appendix B. Field dependency – Hidden Figures Test

The Hidden Figures Test measures how well one can find a simple line figure in a more complex patterns of lines. The user has to select which one of the five presented figures can be found in the pattern (without resizing or rotating it). An example is shown in Fig. B1. As prescribed by [52], the user gets 12 minutes to solve as many of the 16 provided assignments as possible. The field dependency score is the average of two runs with different assignments and is calculated as the total number of wrong and empty answers, expressed as a percentage.

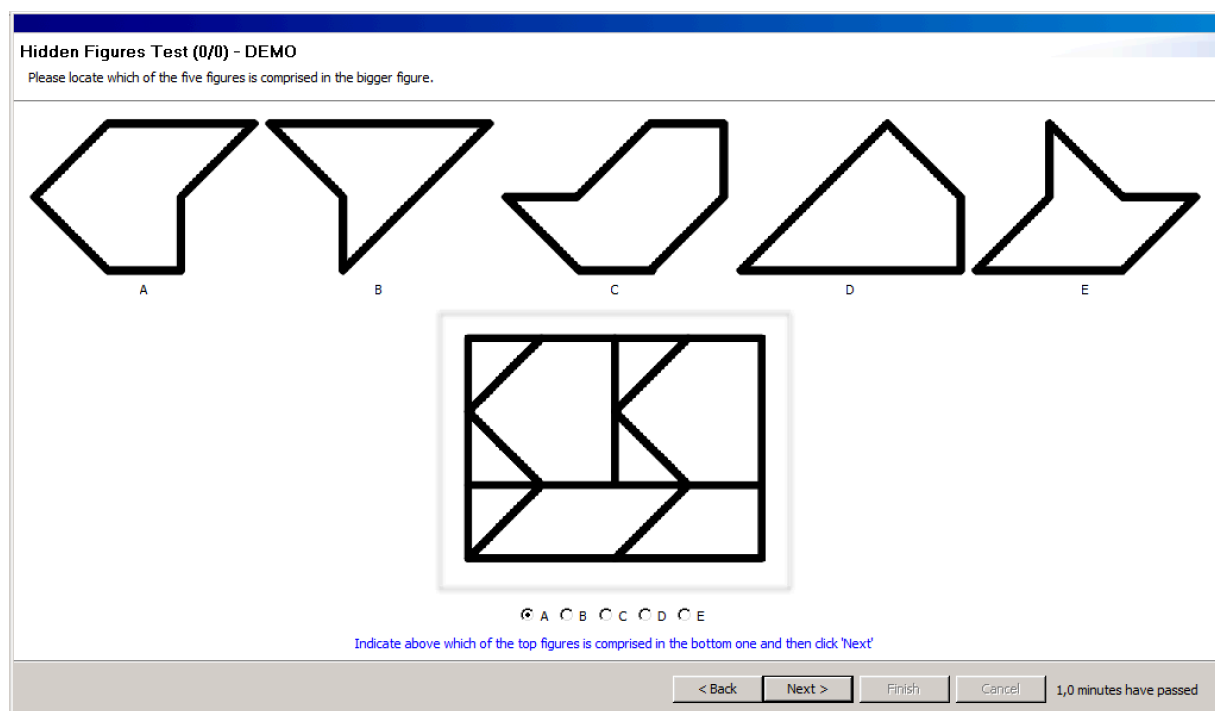


Fig. B1. Example of a Hidden Figures Test assignment.

Appendix C. Need for structure – Questionnaire

The following questionnaire is used to measure the need for structure. It follows the directives of Neuberg, et. al [46]. For each statement the respondent needs to indicate if (s)he (i) strongly agrees, (ii) moderately agrees, (iii) slightly agrees, (iv) slightly disagrees, (v) moderately disagrees, (vi) strongly disagrees (i.e., 6-point Likert scale). The answers on the statements with a minus sign at the end, need to be reversed. Further, it is also indicated to which factor (nfs-1 or nfs-2) the statement belongs. Nfs-1 is the desire for structure, whereas nfs-2 is the reaction to missing structure. Whereas both factors are used to assess the fit of an applied process modeling factor (i.e., for the evaluation of the method), only the desire for structure component is required to perform the steps of the method.

In accordance to [46], item 5 is not used. Yet, it is still included in the questionnaire, because the effect of dropping the question is not documented. The score is the average value of the answer on each statement, which is an integer number ranging from 1 (strongly agree) to 6 (strongly disagree).

1. It upsets me to go into a situation without knowing what I can expect from it. (-, nfs-2)
2. I'm not bothered by things that interrupt my daily routine. (+, nfs-2)
3. I enjoy being spontaneous. (+, nfs-1)
4. I find that a well-ordered life with regular hours makes my life boring. (+, nfs-1)
5. I find that a consistent routine enables me to enjoy life more. (-, /)
6. I enjoy having a clear and structured mode of life. (-, nfs-1)
7. I like to have a place for everything and everything in its place. (-, nfs-2)
8. I don't like situations that are uncertain. (-, nfs-2)
9. I hate to change my plans at the last minute. (-, nfs-2)
10. I hate to be with people who are unpredictable. (-, nfs-1)
11. I enjoy the excitement of being in unpredictable situations. (+, nfs-2)
12. I become uncomfortable when the rules in a situation are not clear. (-, nfs-2)

Appendix D. Measurements for the variables of the SPMT

Fig. D.1. presents an overview of the measurements that were utilized to evaluate the contributions of this paper (see Section 4.4 in the paper). Except for the 3 variables that are already calculated for the *execution* of the method (see Appendices A, B and C), 13 more variables were measured to support the *evaluation* of the method. Their measures are described below.

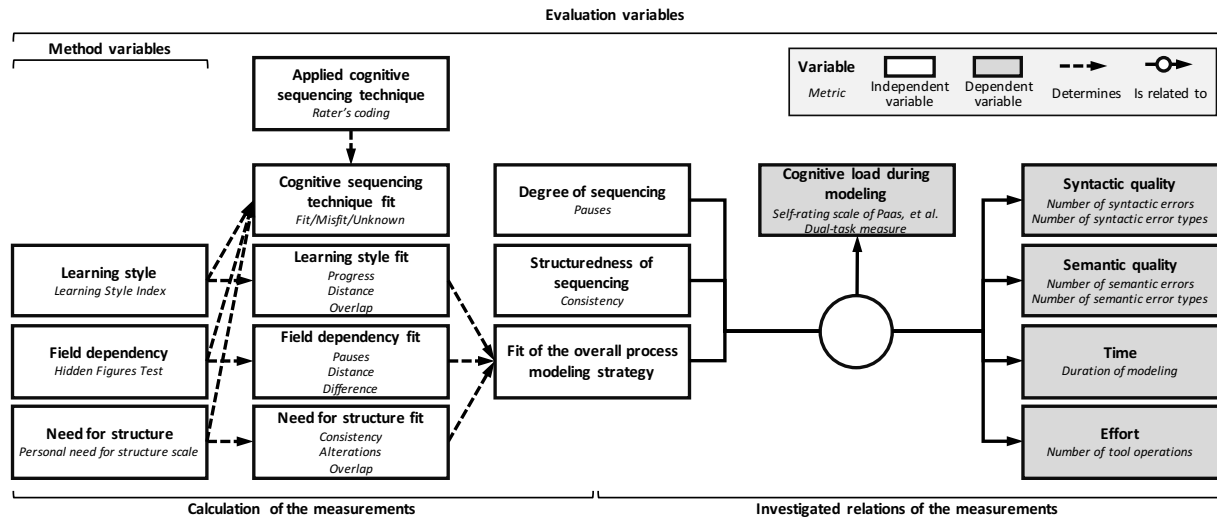


Fig. D.1. Overview of the measurements and their relation

Applied cognitive sequencing technique. For each produced process model a panel of two master and one PhD student independently and in a different order determined which process modeling sequencing technique was applied. Five categories were discovered in previous observations [40,41]: Besides ‘flow-oriented’, ‘aspect-oriented’ and ‘combined process modeling’ (see Section 2.2), a category of clearly ‘chaotic’ modeling was identified and the remainder of the modeling instances was marked ‘unclassified’. The rater’s raw agreement was 57,18%. Because inexperienced modelers rarely apply a sequencing technique very consistently, this agreement value can be considered acceptable. Moreover, for 78% of the modeling instances at least two of the three raters agreed.

Cognitive sequencing technique fit. Then, it was determined if the applied process modeling sequencing technique formed a clear fit, a clear misfit or an unknown fit with the learning style of the modeler. The unknown fit happens for the cases of the category ‘unclassified’, which consist of those that were rated ‘unclassified’ and those where less than two raters agreed on their classification. The

clear misfit was the result of being classified as ‘chaotic’ or of not according to the rules described in the prescriptive derivate of the SPMT (see Section 2.2).

Degree and structuredness of sequencing, fit of the overall process modeling strategy. In order to quantify this degree, structuredness and fit, based on the cognitive literature presented in [47], a set of sequencing and modeling dimensions was first defined.

- **Progress** quantifies the modeling *pace* and is defined as the number of operations per 3 minutes.
- **Distance** quantifies the *locality* of modeling and is defined as the lowest number of elements between the elements upon which two consecutive operations act (following the paths as defined by nodes connected with arcs, regardless the direction of the arcs, including paths through deleted elements if (and only if) necessary, set to the total number of elements if no path between the elements ever existed).
- **Overlap** quantifies the *parallelism* of modeling at the time *t* of a certain recorded operation. It is measured at the time of each operation in the model creation. On that timestamp, the overlap is calculated by looking to all operations of each model element, element per element. For each element, the minimum is taken of the amount of operations before and after the timestamp under consideration. The overlap at a certain time is thus the sum of these minima per model element.
- **Pauses** quantifies the *regularity* of modeling and is defined as the number of occurrences where the duration between two consecutive modeling operations is bigger than the mean plus the standard deviation of durations between all consecutive operations.
- **Difference** quantifies the *order accordance* to the provided case description and is defined as the number of alterations to the modeling operations order to make the order of modeling depth-first for XOR-splits and breadth first for AND-splits (which is the order of the case description, only for elements that are in the final model).
- **Consistency** is defined as the number of raters that agreed on the applied cognitive sequencing technique as described above.
- **Alterations** (of existing elements in the model) is defined as the number of deletion and move operations on any element plus the number of reconnect edge operations.

The degree of cognitive sequencing is thus quantified with the *Pauses* metric. According to the assumption that relatively more structured approaches are easier to recognize, the structuredness of the sequencing is approximated by the *Consistency* metric. The fit of the overall process modeling strategy is derived from the properties related in literature to the cognitive variables (summarized in [47]). The exact formula is presented separately in Appendix E, but the rationale can be found below.

- Sequential learners (in contrast to global learners) should have a steady modeling pace (low standard deviation of *Progress*), model in connected chunks (low average and maximum *Distance*) and model in consecutive blocks (low average and maximum *Overlap*).
- Field-dependent persons (in contrast to field-independent persons) have a short attention span (high amount of *Pauses*), should model in connected chunks (low average and maximum *Distance*), and model in the provided order (low *Difference*).
- Persons with high desire for structure (in contrast to low desire for structure) should apply simple structure-driven processing (high *Consistency*), be confident in their decisions (low amount of *Alterations* and low average and maximum *Overlap*).

Modeling efficacy. In the first place, the modeling efficacy is related to cognitive overload as is proposed by the SPMT. Cognitive overload occurs when the capacity of the working memory does not suffice for handling the necessary information to execute a task. Unfortunately, there are no direct measures for cognitive overload defined in literature. Therefore, cognitive load measures were used as an indicator for cognitive overload. The higher the load, the higher the chance for overload.

Cognitive load during modeling. The emerged standard metric for cognitive load seems to be the self-rating scale of Paas, et al. [63]. This is a validated scale, used successfully in various research designs. Therefore, this metric was adopted to measure the overall cognitive load. In order to measure cognitive load at a more instantaneous level, a dual-task measure for cognitive load was created [64]. Such a metric measures the performance on secondary tasks that occasionally interrupt the primary task. The primary task in this experiment is to create a process model. The most used variant plays a sound to which the user has to react promptly and utilizes the response time as an indication for

instantaneous cognitive load. In order to suit better with the experiment set-up where various participants perform the assignments simultaneously in one room, the recall correctness of a previously shown two-digit number was used as secondary task in the three modeling assignments of the experiment. This number recall task is not an optimal secondary task as it does interfere with the primary task, but it was assumed that the interference was minimal.

Process model quality. For the quality measures an attempt was made to define metrics that distinguish between cognition and knowledge related errors. For this distinction the rational was applied that whenever errors are made consistently throughout the assignment, they are knowledge related and did not count as a cognitive mistake, whereas the others did. Further, both for syntactic and semantic quality, two measures were included: the number of mistakes made and the number of different types of mistakes made. For example, when a modeler used a join gateway with a sign that did not match the sign of the corresponding split gateway twice, this counts for two mistakes, but only for one type of mistakes. The syntactic quality measure was included because it is considered easier to measure objectively, whereas the semantic quality measure was included because it was deemed more relevant (relatively less semantic quality support features exist in today's modeling tools).

Syntactic quality. The syntactic quality was measured by first coding the models, according to the coding scheme in Appendix F. The coding was repeated by three external raters independently and in a different order, which was used to complement and improve our initial coding. The coding system was set up such that it is easy to distinguish between consistent and inconsistent errors, to count the overall number of mistakes and to count the number of different mistake types (see Appendix F).

Semantic quality. A similar coding system was used for the semantic quality (see Appendix G). In order to determine semantic errors, the textual description that the participants had received as input for their modeling assignment was divided in text blocks, each describing one semantic chunk (i.e., a single activity, choice, event, etc.) Next, the text blocks were compared to the model to identify missing, wrong or obsolete model parts, which were coded as semantic errors by two external raters.

Modeling time and effort. The modeling time was defined as the duration from which the modelers started reading the case description towards the last operation recorded in the modeling tool. The modeling effort was measured as the number of modeling operations in the tool (i.e., creation, movement, and deletion operations on model elements).

Appendix E. Measurement of process modeling strategy fit

The quantitative measurement of the process modeling strategy fit is explained below. Different indicators are combined in a single metric, weighted according to their importance for the user.

$$fit = \sqrt[|ls|+|fd|+|nfs|]}{lsfit^{|ls|} \times fdfit^{|fd|} \times nfsfit^{|nfs|}} \quad [1]$$

First, the weights are discussed. Instead of a weighted average, a weighted product was used to minimize the influence of outlier values. It was noted how asymmetric advice is attached to the cognitive variables (e.g., sequential learners are advised to model with a steady pace, whereas global learners are not advised to model with an irregular pace). Therefore, the more a modeler scores to the one side, the more the corresponding advice has to be taken under consideration. Thus the value of the cognitive variables is used as an importance indicator: the weights in the formula.

Next, each indicator in the formula is presented below. They correspond to the instructions discussed in Section 2.2 in the paper. The function $s()$ represents the standard deviation, the function $a()$ represents the average and the function $m()$ represents the maximum of the values of the factor for each modeling operation. The functions $os()$, $oa()$ and $om()$ represent the opposites of these functions (i.e., $1-s()$, $1-a()$ and $1-m()$).

$$lsfit = \sqrt[5]{os|Progress| \times oa|Distance| \times om|Distance| \times oa|Overlap| \times om|Overlap|} \quad [2]$$

$$fdfit = \sqrt[4]{|Pauses| \times a|Distance| \times m|Distance| \times |Difference|} \quad [3]$$

$$nfsfit = \sqrt[4]{|Consistency| \times o|Alterations| \times oa|Overlap| \times om|Overlap|} \quad [4]$$

As indicated by the vertical lines ($|factor|$), all variables are normalized. This means they are converted to a real value between 0 and 1.

In order to improve the accuracy of formula [1], two additional indicator factors were added.

$$lsbestfit = 1 - \frac{abs(bestls-ls)}{23} \quad [5]$$

The factor *bestls* in formula [5] is the learning style score that fits best with the real applied sequencing technique. The factor *lsbestfit* is included in [1] with a weight of 1.

$$nfs2fit = \frac{Structuredness+4}{7} \quad [6]$$

The factor *Structuredness* is the number of raters that coded the sequencing technique as FO, AO or C minus the number of raters that used code UD. The *nfs2fit* factor is included in [1] with a weight equal to $|nfs2|$.

Appendix F. Syntactic errors coding scheme

Below is the coding scheme used by the raters to measure the number of syntactic errors in a model. The complete assignment description used by the raters that performed the coding can be downloaded from <http://www.janclaes.info/papers/SPMM>.

Syntactical error	Code
Contains no end event (but does contain a start event)	0 (0 end events)
Contains no start event	0s (0 start events)
Contains a start event in the middle (ingoing edge on start event)	B (between)
Contains an end event in the middle (outgoing edge on end event)	B (between)
There are multiple end events	E (multiple ends)
One, but not all of the paths are not closed (missing end event?)	P (path not closed)
Contains no split gateways at all	S (no splits)
Forgot some, but not all split gateways	F (forget some)
Contains no join gateways at all	J (no joins)
Contains and, but no xor join gateways at all	Jxor (no xor joins)
Contains xor, but no end gateways at all	Jand (no and joins)
Forgot some, but not all join gateways	G (forget some)
Forgot xor join gateway	Gxor (forget some)
Forgot and join gateway	Gand (forget some)
Forgot xor gateway at end event	Ge (forget some)
One gateway combines a join and split feature	C (combination)
Wrong type of join combined with a certain split	W (wrong type)
Gateway with only one ingoing and one outgoing edge	1 (1 edge in/out)
Wrong nesting of gateways	N (wrong nesting)
AND and XOR are joined together in one join gateway	T (joined together)
Forgot join gateways in case of iterations (edges that go back in the model)	I (wrong iterations)
Some edges are missing (space between two items, but no edge)	M (missing edges)
Used a start event instead of a gateway	GWs (gateway start event)
Used start event as a data object symbol	DB (database)
Used start event where some input is expected (as a message event)	IN (input)
Modeled Petri-net style using start events as places	PN (Petri-Net)
Used empty activities	- (empty)
Other	> (describe)

Appendix G. Semantic errors coding scheme

Below is the coding scheme used by the raters to measure the number of semantic errors in a model. The complete assignment description used by the raters that performed the coding can be downloaded from <http://www.janclaes.info/papers/SPMM>.

Semantic error	Code	Explanation
Obsolete activity	OA	Activity in model that should <u>clearly</u> not be there because it was not described as a separate step.
Obsolete gateway	OGW	Gateway in model that should <u>clearly</u> not be there because it was not explicitly described in the case description.
Obsolete edge	OE	Edge in model that should <u>clearly</u> not be there because the activities it connects should not follow each other (directly).
Obsolete end event	OEE	End event in model that should <u>clearly</u> not be there because the case was not described to end at that place.
Missing activity	MA	An activity is missing where the case description <u>clearly</u> describes something as a separate task.
Missing gateway	MGW	A gateway is missing where the case description <u>clearly</u> describes an optional or parallel split.
Missing edge	ME	An edge is missing where the case description <u>clearly</u> describes that two activities should follow each other directly.
Missing activity because of misplaced iterative edge	MAE	Some activities are not executed in an iteration because the iterative edge points to the wrong starting point of the iteration.
Missing end event	MEE	An end event is missing where the case description <u>clearly</u> describes that the process can end there.
Missing information	MI	Information is missing because a small part of the model is de-scribed by a single activity instead of a more extensive construction.
Wrong condition	WC	The wrong condition is used to indicate which path is executed at what condition after an XOR split.
Incorrect order	IO	A wrongly placed edge causes the flow to be executed in a incorrect order.

Appendix H. Interview protocol

Below is an overview of the interview protocol used as a final part in the pre-test of the experiment. Gradually more information was revealed to the interviewee and questions were asked about the theory and the individual findings of the observational sessions of the pre-test.

1 Modeling styles

SHOW PPMCHART STYLES

- which style do you think you applied
- for each style (1) how well does it fit to you
(2) how difficult do you find it to apply

SHOW PPMCHART OF MODELER

- again, which style do you think you applied
- at a scale from 0% to 100% how consistent did you apply the scale?

SHOW STYLE ON PPMCHART OF MODELER

- do you recognize your style?
- is this a style you apply usually (when writing text or program code)?
- do you feel this style fits to you?
- why (not)?
- is this because you learnt the style or does it feel natural to you?

2 Cognitive load

SHOW PROCESS MODEL

- did you find the exercise hard?
- what was hard?
- can you point out in the model or chart where you encountered difficulties?
- why did it become difficult at that point?
- is this usually why you find things difficult?
- did it go better afterwards?
- when?
- do you think you did well?

SHOW MISTAKES

- why did you make each mistake?

3 Cognitive profile

SHOW COGNITIVE CHARTS

- given your cognitive profile and your adopted approach we expected this/another result, do you agree?
- why does this theory (not) apply to you?
- do you think of situations when this theory might not apply?
- which situations are that?

4 General

- do you have any other questions or remarks about the experiment or interview?

Thank you for your cooperation!

Appendix I. User perceptions

This appendix presents a number of quotes to illustrate the findings presented in Section 4.5.3.

Motivational aspects of the cognitive tests. Participants indicated that *“it was a challenging, fun experiment”*, it was *“sometimes hard, sometimes easy”*, *“it was quite enjoyable”*, they *“really liked this experiment, because it was fun and challenging”*, they found *“the test quite hard, but enjoyed filling it out”*, this was a *“very nice test as preparation for solicitation [=job applications]”* and they found it *“very difficult and I’m seriously starting to doubt my own intelligence”*.

Technical aspects of the cognitive tests. The vast majority of technical comments mentioned *“no interruptions, no technical problems”*, but some participants indicated *“some disruptions during the test”* for example *“coming from two warnings of my anti-virus program”* or *“because my brother came in to ask a question, ignoring the ‘Do not disturb’ sign I added at my door”*. Most disruptions seem to have been minimal and quickly resolved. *“During the second series, the VPN server disconnected. When the problem was solved, I went further with the test without any problems, and I did not need to restart. I could go on where I had left.”* For some students it was not clear how much the disturbance has influenced the results and they were left out from further analysis. For example, *“I was disturbed by a carnival procession which passed with music. [Google translation from Dutch]”*.

Content-related aspects of the cognitive tests. The tool was perceived as *“clear and worked as it should”*. The instructions were mostly described as *“instructions were clear”*, *“very clear”*, *“crystal clear”*, *“good explained”*; although some participants responded that *“instructions weren’t immediately understood but demo made everything clear”* or *“the purpose of the exercises only became clear after the example, so it’s a good thing that we could first practice the exercise”*, which indicates that it may took some effort for certain students to understand them. One, more concerning comment was that for one participant a *“test on paper would be much more convenient”*.

Slowness of the tutorial. A number of comments were directed to this issue, but the participants seemed to show understanding for this matter. *“The system was very slow; this was not encouraging at*

all. The instructions however were clear and detailed.” - “The tutorial was a bit laggy, but you notified us at the beginning so I was prepared. Otherwise it was well put together.”

Positive feedback about the tutorial. A high number of positive and surprised reactions were received. *“I found the information about my cognitive profile very interesting!” “Especially the explanation about my learning personality I found very interesting. I’m happy I know it now.”* A number of participants recognized some aspects explained by the tutorial, whereas others discovered new insights about their own personality. *“I found it interesting to learn more about myself from modeling point of view. Some of those characteristics I had noticed about myself already” - “The feedback on field dependence and flow-oriented modeling was very interesting and confirmed a kind of intrinsic suspicion of me. [translated]” - “I recognize myself mostly in the points raised. They are certainly points that I will remember in order to work in a better and more efficient manner. [translated]” - “The result that I had expected is that I have a strong desire for structure and I actually sit between sequential and global learning. But the second, field-independent, dimension, I did not expect. [translated]” - “Very insightful experiment. One remark: I thought I was rather a structured person whilst the experiment says the opposite.”*

Negative feedback about the tutorial. Only one comment was received about some unclear explanation. *“I found the suggestions for flow-oriented modeling and field independency a bit contradicting. The former instructed to work completely step by step, while as a field-independent person you should not worry about leaving things for later. [translated]”* Although these instructions do not contradict, the student seems to have been troubled about how these guidelines could be combined.

Positive feedback about the method. Most participants were optimistic about the method. *“I did apply both methods, this is clearly the right way of modeling for me [translated]” - “I found the modeling went easier by taking the tips into account, that were given in the context of my profile. [translated]”*. A number of comments were very specific. *“The new aspect-oriented method really is a good way of modeling for me. Because of it I have a better overview on the entire process and I find it*

easier to put the whole thing in one model. By not always laying out in between, I can focus better on the modeling itself. [translated]” - “I liked the whole experiment. I think I learnt something out of it. As such I will try to apply the tip about field-independent working. [translated]” - “Doing the modelling just by using the aspect-oriented method (and more importantly first only focusing on the activities and not yet on the nodes and decisions) greatly decreased the time and effort I had to use while modelling.” - “I have to say that I was shocked by the influence the recommended method had on my efficiency. I was much more sure than last time. In the end, I had to make some small adaptations here and there, but they became always immediately clear to me. [translated]”.

Negative feedback about the method. A number of participant complained that *“It was very difficult to focus on the modelling while keeping the numbers in mind. In the beginning this worked, but at the end I did forget a lot of numbers.” - “Whenever I was really concentrated I forgot my numbers ... but I didn’t have any technical problems really”.* However, this is normal as cognitive load increases towards the end. There were also two comments that mentioned that *“This is what I already did before, so I didn’t really do anything different [translated]” - “I didn’t really feel as if the tutorial helped me changing my way of modelling, structuring... as it described the way I usually work.”* Indeed, some people automatically apply the best suitable strategy and will not experience any improvement with the developed method.

Perceptions about the experiment assumptions. In the experiment, three modeling tasks had to be performed. Much effort was spent to prepare three cases of comparable complexity. For example, our solution to the cases have a comparable number of nodes (17, 26 and 22 for the practice, benchmark and experiment case respectively), arcs (33, 21 and 28 respectively) and nesting depth of routing constructs (each time 3). It was decided to provide each participant with the cases in the same order. This way, they could not prepare for a future modeling assignment by hearing from other participants about the other cases. This also facilitates between-subject comparisons. But one may worry about the validity of within-subject comparisons if the consecutive assignments have a different complexity. The data did not reveal such structural differences. *“The cases in all experiments were well chosen and interesting.”* If anything, the last case was harder, potentially introducing a bias that

would worsen the results. *“I felt as if this task was harder and took longer than the previous two modelling tasks.”* (This reaction is from a participant in the treatment group who was classified with an unknown fit both before and after treatment).

As mentioned earlier, a rather simple BPMN editor is included in the experimental environment. This simple editor was selected to avoid the expertise-reversal effect [58], which describes an initial decrease in performance when an expert needs to be retrained to use another method/tool than (s)he is used to. On the one hand the participants *“found the modeling tool very user-friendly (even more so than previously used tools like ARIS express)”* with a *“helpful tutorial”*. Because of a lack of comments about the editor (and an abundance of feedback on every other aspect of the experiment and tool), it is thus assumed that the participants did indeed not suffer from the expertise-reversal effect. On the other hand, three students missed a feature such as to *“automatically help a bit to put all things in a straight line”* or *“to name my start-events (I created sub processes)”*.